Cassandra Crossing/ Archivismi: l'organizzazione dei documenti in Internet Archive

(562)—Completiamo la descrizione di come Internet Archive organizza i documenti, e di come il sito permette di utilizzarli

---

Cassandra Crossing/ Archivismi: l'organizzazione dei documenti in Internet Archive
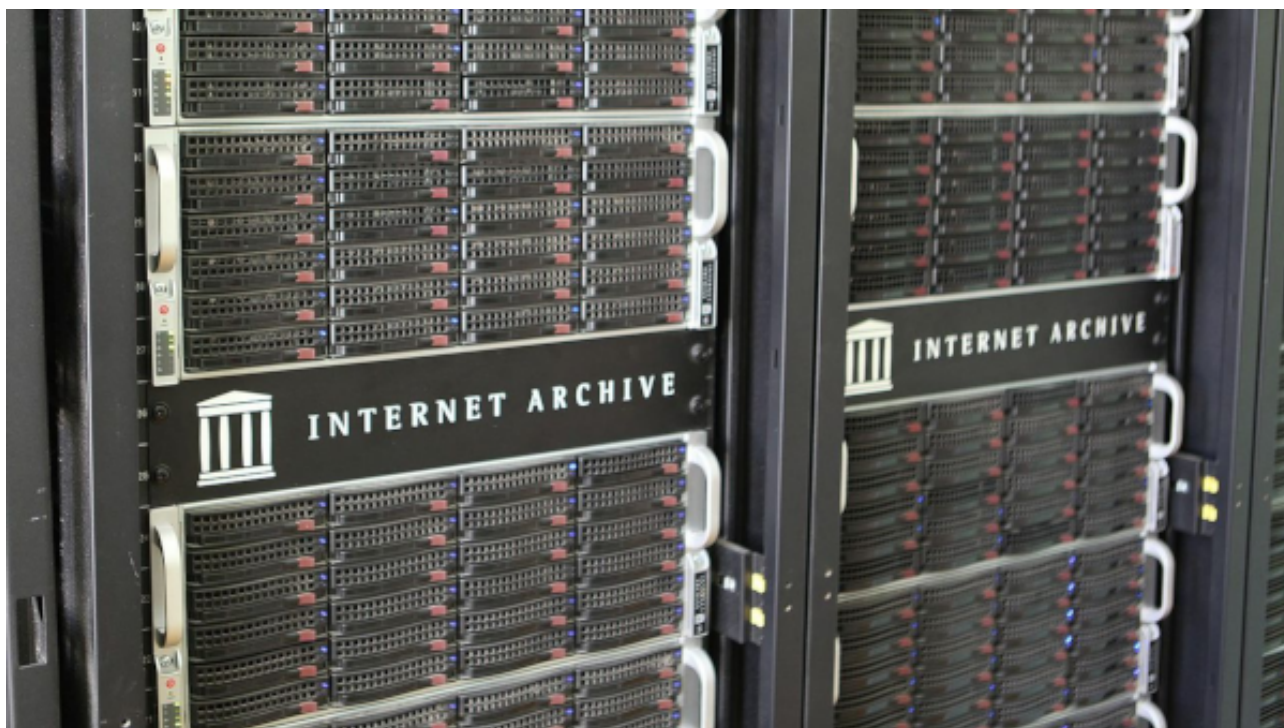
Figure 1:

(562)—Completiamo la descrizione di come Internet Archive organizza i documenti, e di come il sito permette di utilizzarli

28 dicembre 2023—Nella scorsa puntata siamo riusciti ad archiviare documenti ,anche grandi ed in formati eterogenei, e convertendoli durante il processo in modo da averli disponibili in più formati digitali, riutilizzabili per gli scopi più diversi.

Ma per poter dire di aver realmente *archiviato* un documento bisogna anche averlo inserito in un più vasto corpo di documenti, a sua volta dotato di indici e metodi di vario tipo per organizzare e ricercare i documenti e le informazioni in essi contenute.

Facile quindi cogliere l'importanza di sapere *a priori* come una biblioteca digitale già esistente permette di organizzare i propri dati, adeguandosi ad utili e ben studiati standard comuni.

L'architettura di *Internet Archive* è tanto semplice quanto potente.

Il primo livello dell'architettura è l'oggetto, che può essere creato e successivamente modificato in vari modi; un oggetto è tipicamente un singolo documento. Se l'oggetto viene creato da un utente registrato e collegato, all'utente viene assegnato il ruolo di amministratore dell'oggetto, che può quindi modificarlo, arricchirlo di ulteriori file di dati e nuovi metadati,

e così via. Se l'oggetto viene invece creato in forma anonima da un utente non registrato o non collegato, ad esempio utilizzando la Wayback Machine, non può più essere modificato da chi lo ha creato, ma solo dagli amministratori di Internet Archive, dietro specifica richiesta da inoltrare via email, formattata con specifici template.

Il secondo (ed ultimo!) livello di architettura è la collezione (Collection). Una collezione è un oggetto di tipo particolare, formato solo da riferimenti ad altri oggetti. Come tutti gli oggetti è dotato di suoi propri metadati, ma può essere creato solo dagli amministratori di Internet Archive dietro specifica richiesta di un utente registrato, utente che deve possedere certi requisiti, elencati nelle policy di creazione delle collection. Una collezione può contenere altre collezioni come sotto-collezioni. L'utente che si è fatto creare ed assegnare la collezione la può amministrare, inserendoci gli oggetti di cui è il creatore, ad esempio quelli che ha uploadato.

Quando un oggetto viene creato, viene assegnato per default ad una collezione; se l'oggetto è creato in maniera anonima o direttamente da un utente tramite upload, viene assegnato automaticamente ad una collezione che potremmo definire "di sistema".

Ad esempio i documenti che abbiamo creato nelle precedenti puntate, come si può vedere esaminando i metadati nella finestra dell'oggetto o tramite il metadata editor, sono stati assegnati per default alla collezione "*opensource*". Ricorderete che il file dell'articolo usato è stato da noi specificatamente marcato come *oggetto effimero* e destinato ad essere cancellato dopo 30 giorni. Esaminando i suoi metadati, si può notare che è stato assegnato anche alla collezione *test_collection*. Un processo automatico, evidentemente, "spazzola" tutti gli oggetti assegnati a questa collezione e rimuove definitivamente quelli più vecchi di 30 giorni.

Esiste uno pseudo "terzo livello" di organizzazione che è solo di "presentazione", e viene costruito dai creatori del sito assegnando gli oggetti a collezioni particolari ed utilizzandole poi per generare specifiche pagine sul sito di Internet Archive, per favorire un accesso rapido ed estemporaneo a certe categorie di informazioni. Queste sono, ad esempio, le icone che si trovano in home page e sulla barra dei menu del sito.
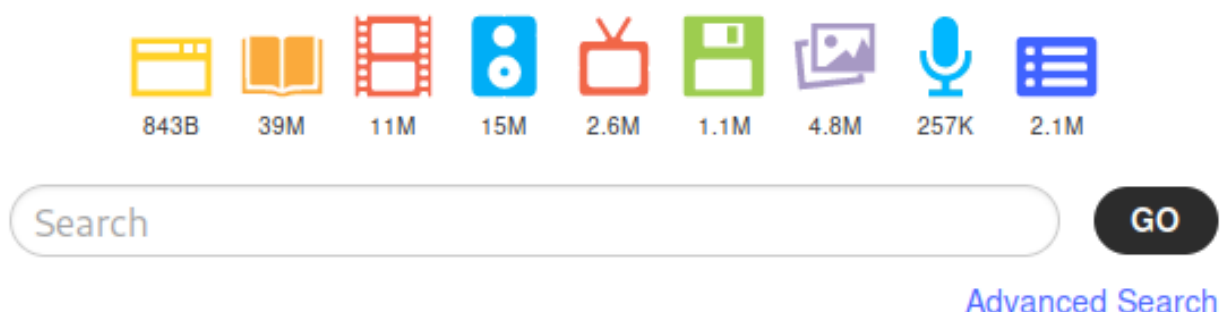
Figure 2:

Il sito di Internet Archive ha un'aria un po' "farraginosa" e retrò. In effetti però, appena

preso un minimo di confidenza, si rivela un meccanismo abbastanza utile e potente per trovare documenti di interesse od avere spunti di cose nuove, che sono di solito collezioni molto accedute.

In realtà, comunque, le informazioni di interesse si trovano, come è facile immaginare trattandosi di una biblioteca, tramite le funzioni di indicizzazione e ricerca, rese disponibili in vari modi sul sito. Ad esempio, visualizzando i propri upload, nella parte sinistra dello schermo si ha accesso ad una serie di categorie di selezione pertinenti, simili a quelle di Amazon.

Figure 3:

Quando necessario, si può accedere direttamente alla funzione di ricerca tramite il box "*Search*" in alto a destra nel sito. Si può accedere alla funzione di ricerca completa cliccando dentro il box stesso e selezionando"*advanced search*".

Figure 4:

Ed anche per oggi è tutto. *Stay tuned* per la prossima puntata di “*Archivismi*”.

Scrivere a Cassandra—Twitter—Mastodon
Videorubrica “Quattro chiacchiere con Cassandra”
Lo Slog (Static Blog) di Cassandra
L’archivio di Cassandra: scuola, formazione e pensiero

By Marco A. L. Calamari on December 29, 2023.

Canonical link

Exported from Medium on August 27, 2025.